

Reference Points for Measuring Social Performance:
A Case Study of a Social Business Venture

Kévin André
ESSEC Business School, France

&

Charles H. Cho
Schulich School of Business, York University, Canada

&

Matias Laine*
University of Tampere, Finland

"Accepted author manuscript" version (postprint).

André, K., Cho, C.H. & Laine, M. (2018). Reference points for measuring social performance: A case study of a social business venture. Journal of Business Venturing, 33(5), 660-678.

Publisher DOI: <https://doi.org/10.1016/j.jbusvent.2017.12.002>

* corresponding author: matias.laine@tuni.fi

Acknowledgments:

The authors remain indebted to the Guest Editors for their extensive support and very helpful comments on earlier versions of this paper. We wish to thank the anonymous reviewers and the participants of the 27th International Congress on Social and Environmental Accounting Research in St Andrews, the 2015 French Congress on Social and Environmental Accounting Research (3rd CSEAR France Conference) in Cergy-Pontoise, the 39th European Accounting Association Annual Congress in Maastricht, the Manuscript Developmental Symposium for the special issue in London, Ontario, and the accounting seminar at the University of Sydney Business School. Charles Cho acknowledges the financial support provided by the ESSEC Research Centre (CERESSEC), the Erivan K. Haub Chair in Business & Sustainability and the Global Research Network program through the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A2A2912421). Matias Laine acknowledges the financial support provided by the Academy of Finland (309884).

Reference Points for Measuring Social Performance: A Case Study of a Social Business Venture

Abstract:

We present a longitudinal qualitative case study to elaborate on how a social venture forms reference points for social performance. Although organizations increasingly use various social performance targets to direct their operations, the scholarly knowledge on social performance reference points remains limited. We make use of the prior accounting literature and draw on the idea of compromising accounts to discuss how provisional and performative metrics can have a significant role in how organizations develop new ways to evaluate their social performance. Given that the social performance reference point criteria are ambiguous and the corresponding referents malleable, performative accounts are helpful as they can intervene in the organizational life by making particular things visible, providing space for interpretations, and facilitating discussion, thus creating temporary settlements and enabling opportunities for productive compromises between different organizational groups and evaluative principles. The recursive feedback loops between reference point referents, criteria and accounting artefacts help the organization to make sense of its own social performance and interpret the associated performance feedback, and thereby provide ground for organizational decisions on further action. Moreover, we discuss how imperfect accounts can be useful for social businesses in their pursuit of developing their activities and achieving social impact.

Keywords:

Social performance; Social entrepreneurship; Compromising accounts; Reference points

Reference Points for Measuring Social Performance: A Case Study of a Social Business Venture

1. INTRODUCTION

In this paper, we present a longitudinal qualitative case study to shed further light on how a social venture forms reference points for social performance (Nason, Bacq and Gras., forthcoming), and how in the organization compromising accounts (Chenhall, Hall and Smith, 2013) are made use of in developing metrics for its operations. We position our work within the established literature on how a firm's strategic behaviour relates to the differences between its goals and performance (Ansoff, 1979; Cyert and March, 1963; Fiegenbaum, Hart and Schendel, 1996). Organizational goals, also known as aspirations or reference points, are used by firms to evaluate their current performance and positions, subsequently leading to their taking different types of actions depending on whether the performance feedback is positive or negative (Gavetti, Greve, Levinthal and Ocasio, 2012; Petkova, Wadhwa, Yao and Jain, 2014). For financial performance, organizations can readily resort to a range of regular, quantifiable and widely used reference points, such as the firm's own historical performance data, benchmarks of peer companies or industries, as well as standard financial statement analysis indicators (Greve, 2002; Petkova et al., 2014). Nonetheless, firms are increasingly utilizing a broader range of reference points to prosper in the marketplace (see Bebbington, Unerman and O'Dwyer, 2014; Nason et al., forthcoming). Recent literature has discussed how firms can either make use of multiple reference points simultaneously (Joseph and Gaba, 2015) or have specific individual reference points outside the traditional financial aspirations, such as social performance reference points (Nason et al., forthcoming). However, the theoretical knowledge of these settings remains limited, and further research has been called for (e.g. Joseph and Gaba, 2015; Shinkle, 2012). This lack of knowledge on social performance reference points and the associated firm behaviour is intriguing, given the rising prominence of social performance metrics and evaluation across markets (Bebbington et al., 2014).

Recently, however, Nason et al. (forthcoming) have proposed a novel theoretical approach to social performance reference points. On a conceptual level, Nason et al. (forthcoming) maintained that in contrast to financial performance, social performance is ambiguous and its reference points malleable. Due to this absence, Nason et al. (forthcoming) argued that firms engage and negotiate them together with their firm-identified stakeholders (Zavyalova, Pfarrer,

Reger and Hubbard, 2016). They also proposed that both avenues of performance feedback and the corresponding strategic responses undertaken by the firm in relation to social performance reference points differ from those related to financial performance. Furthermore, and of particular relevance for the present paper, Nason et al. (forthcoming) theorized that in terms of strategic responses to performance feedback, social enterprises would differ from non-social enterprises because, in the context of social enterprises, social performance is a key element of identity for both the firm and its firm-identified stakeholders (Battilana, Sengul, Pache and Model, 2015).

While Nason et al. (forthcoming) limited their discussion to the theoretical level, they called for empirical studies to shed further light on the formation, setting and subsequent interpretation of social performance reference points. Accordingly, in this paper we respond to this call by presenting a qualitative case study, the empirical material of which was collected through a longitudinal six-year action research project of a social enterprise to start and operate a water treatment plant in a poor rural region of Bangladesh. In this paper, we discuss two questions. First, we ask how social performance reference points were formed during the early stages of the social venture. We addressed this question through a longitudinal case study because it allowed us to provide a dynamic view of how an organization goes through several reference points to make sense of how its own performance should be evaluated. Second, we discuss how these social performance reference points were interpreted and enacted by different organizational actors below the board level. As most of the prior work on reference points focused on firm-level aspirations (e.g. Shinkle, 2012), we maintain that further insights can be provided by going inside the organization and exploring how key actors at different organizational levels make sense of and act on changing reference points over time.

We argue that our investigation contributes to the literature on social performance reference points in several ways. First, we employ a longitudinal case study to supplement the conceptual framework (Nason et al., forthcoming) with empirical evidence of a real-life setting. Our case study presents evidence on how an organization seeks to find appropriate ways of establishing social performance reference points. At the same time, we answer calls in the recent literature for more in-depth qualitative research to shed light on the internal organizational processes and realities in the context of reference points (e.g. Holmes et al., 2011; Shinkle, 2012). Second, our detailed empirical material allows us to move beyond firm-level considerations and to discuss how the introduction of new social performance reference points were received, interpreted and subsequently engaged within the case organization. Several fields of enquiry

have shown that middle managers and other organizational actors play a significant role in how decisions made at the upper echelons of a firm are eventually implemented (e.g. Balogun and Johnson, 2005; Busco and Quattrone, 2015; Hyvönen, Järvinen and Pellinen, 2008). Here, we make use of the prior accounting literature and the concept of compromising accounts developed by Chenhall et al. (2013; see also Stark, 2009) to discuss how the employees within the case organization made sense of the new reference points and sought to create metrics for evaluating the organization's performance in relation to the changing reference points. Third, we discuss how over a period of several years, our case organization experimented several social performance reference points and drew on its past experiences during each round. We show how in this context provisional and performative metrics (Dambrin and Robson, 2011; Vollmer, Mennicken and Preda, 2009) have a significant role in the formation of social performance reference points (Nason et al., forthcoming), as they intervene in organizational life by making particular things visible, mediating between different interests and facilitating discussion, thus creating temporary settlements and enabling productive compromises between organizational groups holding different evaluative principles. Finally, our case study adds to discussions on the challenges of measuring social performance in social enterprises. As the idea of social entrepreneurship has become more popular and the number of initiatives in the field has grown significantly, social enterprises are increasingly under high pressure and have an increasing number of demands to demonstrate the impact of their activities (e.g. Carnochan et al., 2013; Nicholls, 2009). In practice, this has proven complicated, and there is an ongoing debate over how the social performance of such organizations should be evaluated (Bellucci, Bagnoli, Biggeri and Rinaldi, 2012; Mäkelä, Gibbon and Costa, 2017). Despite calls for common frames of evaluation and a plethora of available tools (see Anner, 2016; Mook, Chan and Kershaw, 2015), the recent literature has highlighted how individual social entrepreneurs tend to ignore ready-made alternatives and instead use bricolages of context-specific metrics to build their own measurement frameworks (Molecke and Pinkse, 2017). Our investigation produces further empirical insights into the early stages of social venture emergence (Perrini, Vurro and Costanzo, 2010), including how a case organization comes to use particular reference points to evaluate its social performance and how, in this process, individual actors make sense of multiple competing reference points.

This paper includes five sections. The second section reviews the prior literature and discusses the theoretical concepts we make use of in this paper. The third section describes the case study

as well as the data and methods used. The findings are presented in the fourth section. The fifth and final section discusses and concludes this paper.

2. LITERATURE REVIEW AND THEORETICAL FOUNDATIONS

2.1 Reference Points

In reviewing the prior literature on organizational aspirations, Shinkle (2012) discussed three distinct theoretical frameworks: behavioural theory, Ansoff's strategic management view and strategic reference point theory. Shinkle (2012) pointed out that despite having different backgrounds, assumptions and mechanisms, all three frameworks are based on a similar logic in which a firm compares its actual performance to a desired performance level, known as an aspiration, goal or reference point. A firm is considered to alter its strategy and take actions depending on how it performs relative to its reference point. As long as firm performance is above the reference point and, hence, the firm is achieving its aim, firm behaviour is argued to remain relatively unchanged and to become increasingly risk-averse. Alternatively, should a firm's performance be lower than its goal, then the firm is expected to engage in problemistic search and to explore new strategies to enhance its performance (e.g. Cyert and March, 1963; Kuusela, Keil and Maula, 2017; Petkova et al., 2014).

Firm aspirations have long been theorized to consist of established financial reference points against which firm performance is evaluated. For instance, these reference points have been based on historical comparisons to the firm's past performance, or on social comparisons within peer industry groups or to other similar organizations (Moliterno, Beck, Beckman and Meyer, 2014). Organizations have also utilized multiple reference points to assess their performance, although the theoretical understanding of how exactly this takes place is still limited (Hu, Blettner and Bettis, 2011; Joseph and Gaba, 2015). Still, the scholarly work on multiple reference points has often focused solely on financial aspirations, even though it is well established that in contemporary societies, organizations increasingly have goals and targets with criteria beyond traditional financial metrics (Bebbington et al., 2014; Ebrahim and Rangan, 2014). Such goals are less institutionalized and, hence, there is substantial variation across industries, geographical contexts and organizations (Luo, Wang, Raithel and Zheng, 2015). Organizations strive to reach both environmental and social goals, from greenhouse gas emissions and wastewater effluents to employee welfare and social impact (see Bebbington et

al., 2014). Despite the increasing significance and prevalence of alternative criteria, there is still limited scholarly knowledge on how organizations decide and set these reference points, how they are prioritized in relation to financial goals, and how social reference points are understood, interpreted and subsequently enacted within organizations (Nason et al., forthcoming; Shinkle, 2012). These questions are significant because the selection of reference points is argued to signify organizational priorities and overall direction and, hence, to have implications for organizational choices (Petkova et al., 2014). Moreover, Fiegenbaum et al. (1996) argued that in addition to an organization's strategic choices, top management can also influence organizational performance through the choice of particular reference points.

Answering frequent calls in the literature (e.g. Holmes et al., 2011; Joseph and Gaba, 2015; Shinkle, 2012), Nason et al. (forthcoming) presented a theoretical framework for social performance reference points, which they argued differs from financial performance reference points in several ways. First, in terms of feedback formation, Nason et al. (forthcoming) maintained that while firms select their financial performance reference points from a broad range of established metrics (Luo et al., 2015), no established benchmarks are used as reference points for social performance nor is there consensus on how firm social performance can be measured (see Shinkle, 2012). Second, the feedback about performance is framed differently. For financial performance reference points, negative feedback is expected to engage firms in problemistic search to find ways to enhance performance, while positive feedback is expected to trigger strategic conservatism to maintain their current positions (Schillebeeckx, Chaturvedi, George and King, 2016). For social performance reference points, Nason et al. (forthcoming) proposed that negative feedback triggers a legitimacy frame, meaning that the firm tries to avoid losing key stakeholder support as a result of not meeting stakeholder expectations. Positive feedback, on the other hand, triggers an efficiency frame related to whether the firm puts too much emphasis on social performance, the financial implications of which are uncertain (see Margolis and Walsh, 2003). While theories of financial reference points expect firm reactions to be stronger when firm performance is relatively close to the reference point (Petkova et al., 2014), Nason et al. (forthcoming) argued that when it comes to social performance reference points, a firm takes stronger actions when the gap between the reference point and actual performance is wider, and a firm takes little action when the performance gap is smaller. Third, Nason et al. (forthcoming) proposed that firm identity impacts how firms respond to social performance reference point feedback. They argued that if social performance is a central part of the organization's identity, that organization reacts differently to social performance

feedback. Specifically, Nason et al. (forthcoming) gave the example of social enterprises as entities for which social performance is a central part of their organizational identity.

As most of the prior theoretical discussion on reference points has focused on firm-level aspirations, calls have been made to explore how managers make sense of various reference points, particularly in settings where these reference points and their related feedback are ambiguous (see Aranda, Arellano and Davila, 2017; Joseph and Gaba, 2015). Shinkle (2012), for instance, pointed out that in investigating reference points and performance level, scholars should examine the differences in viewpoints between top management and those in other levels of the organization. Likewise, Nason et al. (forthcoming) suggested that further research look into how individuals receive, make sense of and assess both reference points and social performance feedback (see also Joseph and Gaba, 2015). This resonates with prior work in strategy and strategic management that has highlighted the role of middle managers and other organizational actors in implementing strategic changes (Balogun and Johnson, 2005). In particular, Balogun and Johnson (2005) emphasized the relevance of understanding the lateral and internal processes through which organizational actors make sense of top-down strategic change initiatives, which subsequently affect the outcomes of such policy initiatives. Similarly, accounting scholars have long acknowledged the complexity of organizational processes in relation to accounting change and highlighted that it is important to investigate how organizational key actors make sense of and enact such changes (see Ahrens and Chapman, 2006; Briers and Chua, 2001).

For this purpose, we draw on recent work by Chenhall et al. (2013), whose work on compromising accounts provided the conceptual tools to discuss how organizational members develop tools to evaluate social performance in response to changing reference points.

2.2 Compromising Accounts

The idea of performance measurement tends to be associated with striving for a nuanced accounting artefact (Englund, Gerdin and Burns, 2011), that is a formal or material system, which provides precise and objective knowledge to aid organizational decision-making (see Burchell et al., 1980; Miller and Power, 2013). However, imperfect accounts can also be helpful to an organization through, for instance, bringing organizational actors together and developing a common way of seeing, which may result in constructive processes that aid organizational functions (see Busco and Quattrone, 2015; Carlsson-Wall, Kraus and Messner, 2016). In this context, Chenhall et al. (2013) drew on Stark's (2009) idea of organizing dissonance and

developed arguments for how particular kinds of accounts can become compromising accounts, which can be productive for organizations in settings with tensions between different evaluative principles (see Carlsson-Wall et al., 2016; Pache and Santos, 2013, 2010). Chenhall et al. (2013) drew on a case study of a non-governmental organization (NGO) in which some internal stakeholder groups held conflicting views on the development of the organization and the prioritization of certain actions. Chenhall et al. (2013) maintained that for a particular account or performance measurement system to work as a compromising account and to create productive friction, the account or system should fulfil some of the expectations of each relevant, conflicting evaluation logic. In this way, such accounts can create concurrent visibility for competing demands and logics, and thereby create space for positive debate and dialogue, potentially creating productive friction that helps to reconcile conflicting demands or expectations. It is noteworthy that compromising accounts can feature both as a process, which “lead to and follow up from the physical production of a material account”, and as a noun, signifying “the account itself that is produced in some material form” (Chenhall et al., 2013, p. 281).

More specifically, Chenhall et al. (2013) pointed out three essential processes by which compromising accounts become productive for an organization. First, the accounts should be imperfect when enacted in practice. Such imperfection facilitates the creation of productive friction, as it requires organizational actors with different views, values and expectations engage in discussion and debate (see Stark, 2009). Second, the account or performance measurement system must give space and visibility to features important to different groups (Carlsson-Wall et al., 2016). This implies that both in designing and using the accounts, different evaluative principles should be visible simultaneously, as this concurrent visibility enhances the productivity of compromising accounts (Chenhall et al., 2013) through, for instance, allowing those using the accounts to see the organization from different perspectives (Goretzki and Messner, 2016). It is worth noting, however, that too much concurrent visibility may also result in a cumbersome system, with a great number of different indicators speaking to the demands of various groups. There is danger that opportunities for dialogue and productive friction may become stifled by the sheer number of indicators. Third, and finally, Chenhall et al. (2013) noted that the productiveness of compromising accounts hinges on the extent to which the debate and discussion focus on the evaluative principles underlying the accounts, as opposed to the technical details of the accounting measures. The authors suggested that by focusing more on the evaluative principles, the actors can better engage with latent conflicts and thereby

proceed with productive friction that initiates discussions about how different evaluative principles could be integrated (see also Kraus, Kennergren and von Unge, forthcoming).

Similar observations have been presented by others. According to Busco and Quattrone (forthcoming, p. 2), different forms of accounts and visualizations can ‘construct shared meanings and platforms of mediation to stabilize and mediate among diverse interests’ (Busco and Quattrone (forthcoming, p. 2; see Briers and Chua, 2001). Perfection is not a necessity, as incomplete information can also have a positive impact on organizational dynamics and managerial action. At the same time, however, consensus and compromise are often not reached and, hence, accounting systems are used in organizational settings without ‘agreement, alignment and transparency’ (Busco and Quattrone, forthcoming, p. 2). The complexity and institutionalization of the utilized accounting system may be relevant. Kilfoyle et al. (2013, p. 390) argued that vernacular accounting systems, i.e. those that are more mundane and less formal, are ‘more likely to retain compromises or to create pragmatic hybrids’, because the information included lacks purity and retains ‘the complexities of organizational values and rationalities’ in the system. Such observations resonate with the complexity often present in a social enterprise in which inherent tensions often arise between different evaluative principles, manifesting in competing social and economic goals (Pache and Santos, 2010).

Accordingly, in our empirical investigation, we draw on this body of prior accounting literature. In particular, we make use of the conceptual idea of compromising accounts (Chenhall et al., 2013) to analyze how social performance reference points were formed, developed and employed in our case organization, and to discuss how these reference points were interpreted and acted on within the organization.

3. THE CASE ORGANIZATION, DATA AND METHODS

3.1 The Case Organization

As discussed above, vast prior literature has drawn on the behavioural theory of the firm, and the relationship between firm aspirations and subsequent performance is well established. Most of this literature, however, is based on archival methods; as such, many recent calls have been made for more in-depth case studies that include direct engagement with individual managers (e.g. Holmes et al., 2011; Joseph and Gaba, 2015; Shinkle, 2012). Accordingly, we present a single-case study (Patton, 2002; Yin, 2009), which we argue is useful for making sense of ‘how

and why some social phenomenon works' (Yin, 2009, p. 4). As the scholarly knowledge on social performance reference points is still limited (Nason et al., forthcoming), we maintain that a detailed qualitative case study is an effective method for providing insights into this phenomenon in the natural context (Bédard and Gendron, 2004) of a particular organization. Following Gendron (2009), we use a single-case study to provide a thick description of organizational practices (see Ahrens and Chapman, 2006) to shed light on the complex processes of social performance measurement.

The case organization we focus on in this study is a joint venture launched in 2008 by a Bangladeshi non-governmental organization (NGO) and a French multinational corporation (MNC). The case organization operates a water treatment plant that provides drinking water to more than 6,000 people in rural Bangladesh through public tap points and private house connections. The organization operates in a region where the groundwater resources are, to a large extent, contaminated with arsenic. Such arsenic-contaminated water is a significant health concern within Bangladesh: arsenic does not impact the smell or taste of the water, but it has substantial adverse long-term health consequences. It has been estimated that, in Bangladesh, some 25 to 50% of the population drink arsenic-contaminated water on a daily basis (Argos et al., 2010).

A significant feature of our case organization is that it operates as a social enterprise. Its objective is to provide safe drinking water to people in rural Bangladesh at a price that makes safe water affordable for all. Since its foundation, the mission of the organization has been 'bringing safe drinking water to the people of the country, to the poorest and the disadvantaged people particularly in the remote areas' (Joint Venture Agreement, 2008). However, while the organization focuses on targeting the very poor segments of the population, the founding partners have emphasized that it should not operate as a charity. The explicit goal of this social enterprise is to deliver social value, but it is supposed to do so in a financially sustainable manner without any subsidies or grants. As such, it is an example of a hybrid organization embedded in two institutional logics, which implies the pursuit of different and, possibly conflicting, goals (Pache and Santos, 2010).

The case organization selection was based on several criteria. We have had access to the organization since its foundation, which provides us a unique opportunity to understand how the evaluation of social performance has developed within the organization. As such, we maintain that this setting provides us with an extreme case (see Yin, 2009) in the sense that we

can discuss how the use and interpretation of social performance reference points have evolved from non-existence through several iterations. This social venture and the corresponding setting are also particularly suitable for analyzing the development of social performance, because there were no readily available tools or metrics the case organization could have applied; it had to develop them. Furthermore, following Yin (2009; see also Patton, 2002), we argue that selecting a longitudinal case study allows us to understand how things change over time. We consider our study an exploratory case (*ibid*), as our intention is not to explain an empirical phenomenon but rather to shed further light on that which has limited scholarly knowledge.

3.2 Data Collection and Analysis

In our investigation, we draw on a longitudinal action research project, which was based on a cooperation between an academic institution and the case organization. The project lasted for six years, between December 2009 and December 2015. Our active involvement with the organization commenced in March 2011, when the first named author joined the project. It is, however, relevant to point out that while our role was merely observational for most of the events we discuss in our empirical section, our lead author was actively involved in designing the key performance indicators (KPIs) the organization developed and began using towards the end of the discussed timeline. We acknowledge that this intervention needs to be taken into account when interpreting our findings; hence, we discuss it in more detail in our concluding section.

We base our study on two primary datasets. First, we have in our possession a broad archival dataset that consists of some 2,500 documents collected during the project. These documents were produced for various purposes and include, for instance, financial and social reports, sales reports and forecasts, various governance documents, minutes from board meetings and third-party evaluations. The broad categories of the documents are listed in Appendix A. Our investigation does not make a systematic analysis of these documents. Instead, we focus on those we considered to be relevant to our research topic, the formation of social performance reference points. Moreover, the documents were instrumental to our study, as they formed the backdrop for interviews and subsequent analysis.

In addition to the documents, our second primary dataset consists of 36 interviews (see Appendix B) that were conducted in two stages: the first 11 interviews were undertaken in 2011–12, and the latter 25 interviews were conducted mostly in early 2014. In the first set, the interviews were more broad to establish an understanding of the operating logic of the social

venture as well as of the views of the various participants. For the second set of interviews conducted in 2014, we developed the interview guide on the basis of the knowledge we derived from the previously collected material, including internal documents, prior interviews and the action research project. However, it is worth noting that this interview guide was followed loosely and was adapted for each interview, as strictly following a single guide would not have served the purposes of the study, given the interviewees' diverse roles, organizational positions and professional backgrounds. The interviews were mainly undertaken by the first named author. Most of the interviews were conducted in either English or French, in which the authors are fluent, although some interviews undertaken in Bangladesh were held in Bangla with the aid of an interpreter. All the interviews were recorded and have subsequently been transcribed verbatim. The length of the interviews was between 25 min and 3 h; taken together, we have some 34 h of interview material.

Our analysis of the interviews began with a thematic analysis, during which we examined the interview transcripts to gain a broad understanding of the data and the most prominent issues. We proceeded with our analysis by using manual coding of the transcripts to identify those segments and statements in which the interviewees discussed our topics of interest. This data reduction (Miles and Huberman, 1994) allowed us to focus on our key questions of interest: how the interviewees perceived and understood success in the organization, how different evaluative principles were referred to, and how the relationship between organizational performance and societal impact were discussed and understood. At this stage, we engaged in data display (Miles and Huberman, 1994) and used various matrices to organize our observations and interpretations. We paid attention to how individuals at different organizational levels and with varying roles described the organizational goals and their development over time. Moreover, we explored our document-based dataset to find further evidence for our interpretations and to rule out conflicting views and rival interpretations (Patton, 2002). Overall, our data interpretation (Miles and Huberman, 1994) occurred over several rounds, during which we aimed to make sense of our empirical material through various theoretical lenses. The final interpretation was formed through several rounds of data interpretation as well as a continual interplay between the theoretical underpinnings and our empirical material. Eventually, we formed a shared interpretation of how the organization formed its social performance reference points over time and how these were interpreted by the employees at different levels.

In presenting our findings, we focus on the key events and debates that best illustrate how social performance measurement developed within the case organization and how these were made sense of by organizational members. We acknowledge that organizational events and institutional changes do not flow in an orderly and linear way, and stages can seldom be identified in a definite fashion (see Andon, Baxter and Chua, 2007). Instead, such change processes tend to be characterized with uncertainty and serendipity (see Busco and Quattrone, 2015), implying that in an organizational context, the continual flow of unintended and unexpected events affects outcomes. Indeed, a linear narrative focusing on particular events necessarily simplifies the complexity of an organizational reality, often leading to rationalistic post hoc explanations for change. At the same time, however, we maintain that focusing on key episodes is warranted for analytical purposes, as our intention is not to make definite generalizations of how things are but rather to highlight those aspects that should be taken into account in the theoretical development of reference point formation. A similar empirical approach has, for instance, been used in prior work using actor network theory to explore management accounting change in organizational settings (e.g. Hyvönen et al., 2008; see also Briers and Chua, 2001).

Having discussed our theoretical underpinnings as well as the data and methods used in our investigation, we proceed to the presentation of our empirical findings.

4. EMPIRICAL FINDINGS: REFERENCE POINTS IN THE CASE ORGANIZATION

In this section, we track and discuss how our case organization, a social venture, sought to develop the evaluation of its social performance. Our analysis focuses on the case organization's first few years of operation, during which the case organization strove not only to grow and increase its number of customers but also to learn how to evaluate whether it was fulfilling its mission statement. We identify four key reference points against which the organization's performance was compared at different points in time. We then discuss how these reference points were set at the board level and how they were subsequently interpreted by the employees within the organization. Moreover, we present the debates that took place around each reference point and how these compromising accounts helped solve the debates and find imperfect yet productive ways to assess the organization's performance around said reference points.

Our empirical findings led us to the development of a conceptual model, which shows how social performance reference points are formed in an organization (see Figure 1). We find that the setting of the reference points takes place through a process of compromising accounts (Chenhall et al., 2013), in which organizational actors and groups go through discussion seeking ways to combine different evaluative principles they hold (Stark, 2009). In this process, we propose that there is a recursive relationship between reference point criteria, reference point referents and accounting artefacts, whereby each of these informs the others in an ongoing fashion, also providing space for interpretations by organizational groups. Through this debate and dialogue the organization selects on particular metrics, which subsequently will be produced through the accounting artefact. It is here worth noting that the accounting artefacts (Englund et al., 2011) and the metrics they produce are performative, implying that they do not just mirror and reflect the organizational activities, but they also intervene in the organizational reality by affecting how people talk, frame and work around particular issues (Carlsson-Wall et al., 2016; Dambrin and Robson, 2011). The accounting artefacts and the associated metrics create visibility and hence clarify the possible difference between referents and criteria, leading the organization to take appropriate strategic decisions to respond to the feedback. Simultaneously, the performativity implies that the metrics enable new interpretations to be made of the organizational reality, leading to continued discussion and further iteration (see Busco and Quattrone, forthcoming). The metrics are hence only temporary and provisional, and as such subject to recrafting.

It should be noted that all accounting devices and associated metrics are not performative in a similar way. This is illustrated in our empirical setting, in which the performativity of the metrics increased over time: While in the early stages the production of metrics was more sequential with only limited feedback loops, in the later stages the metrics became increasingly interlinked with and internal to the process of compromising account, and have thereby stronger recursive interplay with the reference point referents and criteria. We will now move on to discuss our empirical findings in more detail.

4.1 Reference Point 1 (RF1): Health Impact

Since the launch of the case organization, the board discussed whether and how it would be possible to assess the social impact of its activities. The organization was designed to operate as a social business and, as such, its key mission was to improve the long-term health of the

villagers in the area by offering them water that would reduce their exposure to arsenic-contaminated water. On the board level, everyone agreed on the mission, but there were different views as to whether it would be worthwhile to assess that impact from the very beginning. MNC representatives argued that to evaluate the actual impact in the long term, it was essential to know what the situation had been before they started to operate. However, not everyone was convinced that such a health impact study was worth the effort:

We would not have been able to speak about the positive impact without measuring this. But [the chairman of the NGO] was not willing to conduct a study or to develop tools [for it]. So when we first spoke about this, we agreed on the business plan and then we [at MNC] decided to conduct a sanitary study, which would measure whether and how people were contaminated with arsenic. The goal was to conduct another sanitary study a few years later to compare it with the situation we were facing at the beginning. So it was about health, and [the chairman of the NGO] said ‘okay, it’s your budget and you can use it the way you want’. (Project manager in Paris, 2014)

Despite the controversies, the board decided to go ahead with the first stage of the health impact study before launching the water processing plant in April 2009. The case organization did not have the necessary expertise for the study, so the study was conducted jointly with MNC’s research centre and an independent Bangladeshi research centre (RC). Moreover, involving external experts was important because it increased the credibility of the results.

This health impact study was structured as a comparative survey and was inspired by the experimental methodologies recommended by economists such as Duflo (2010). The aim was to provide scientific evidence of the impacts by comparing the outcomes for a group of beneficiaries with those of a control group who were not targeted by the activities. Such randomized control trials are widely used by NGOs and public agencies, and MNC also had prior experience using them in another context (Devoto et al., 2012). Accordingly, the ex ante survey was carried out in April 2009 to gather data on both the villagers’ water use and the arsenic exposure in the area, while the ex post study was planned to include a sub-sample of the survey to allow for comparisons and, hence, assessment of the social impact of the activities.

The ex ante study showed a correlation between the presence of arsenic in the drinking water sources and the arsenic contamination of the villagers. Approximately 40% of the drinking water sources in the area were contaminated with arsenic, and 99.8% of the residents were found to have increased levels of arsenic in their urine. At the same time, however, it was found

that the villagers used different water sources for different purposes, and those used for cooking were found to be arsenic-free. On the whole, the survey strengthened the case that there was a need for pure drinking water in the area.

The results of the health impact survey set the first reference point for the organization's social performance. It quickly became evident, however, that this was insufficient in practice, as the metric only showed that the project could be relevant, but it had no relation with the organization's activities (see Figure 1). Moreover, after the first few months of operations, the board realized that the organization was selling substantially less water than expected, even though the organization had extended its operations and the quality of its water was excellent. The low level of consumption also implied that they would not be able to identify any statistically significant findings on the health impacts of the project. Hence, before any formal feedback on the health impact was received, the health impact survey and its ex post phase were set aside and the board started looking into new ways to assess the organization's social performance.

4.2 Reference Point 2 (RF2): Water Sales

At this stage, it became evident that several of the case organization's initial assumptions in the business plan were not holding true:

We started from a business model that was based on two assumptions: potential consumers would use the water for drinking and cooking, up to 30L/day/household, and we would have a penetration rate of around 70% in that area. When we began the project, we quickly realized that the penetration rate was 10%—it has increased since but it was 10% at the beginning. And people—as always—consumed the water only for drinking. So instead of 0.70 times 30, it was 0.10 times 10. It was very far from the reality in terms of economic performance. (Case organization's Managing Director, 2011)

At the same time, the focus shifted from external social impact to internal performance. The case organization was already under substantial external pressure, as both MNC and the NGO had communicated the launch of the joint venture broadly and the perceived limited results were raising questions. For instance, the head of sustainability of MNC, who did not take part in the project but was nonetheless very keen to monitor its progress, expressed his concern over the situation:

Questions were raised when we found out that the business plan was far from being reached. It was not a secret because [the chairman of the board] had said that there was a concern. (MNC's head of sustainability, 2011)

Based on its own consideration and also as a response to these stakeholder expectations, the board decided that the key priority would become to increase water sales and not to prove the health impact of its water. Hence, a focus on social impact assessment would not have improved organizational management, as the internal processes would have remained a 'black box' (Barrett and Carter, 2010; Mulgan, 2010). Thereby, the evolution of sales became the new reference point for the organization's social performance.

The level of water sales in the targeted rural area served as the new criterion (Nason et al., forthcoming), which was initially compared to the sales forecasts in the business plan based on the initial assumptions about penetration rate and average consumption. As the difference between the forecasts and the actual performance was major, the performance feedback remained very negative. Hence, it was soon decided that it would be better to compare the sales with previous months, which would allow the performance evaluation to proceed in a more positive light. In other words, there was recursive iteration in the search for an appropriate referent and suitable criterion.

Yet, several employees did not perceive water sales to be an appropriate reference point for social performance and held that the previous reference point based on health impact should also be considered. Various reasons were pointed out for this, especially by the international volunteers we interviewed. First, a figure of water sales does not guarantee that the villagers were using it as their drinking water: the consumption could have been for purposes that would have no impact on arsenic contamination and, hence, no social impact. Second, in the absence of scientific evidence on the arsenic contamination of all available water sources in the area, the organization could not claim to be the only safe water source for the villagers:

In fact, the only thing we could tell them—and we did—was that if they wanted water that was 100% safe and that was not contaminated with arsenic or uranium or something else, then they had to take our water because we actually tested the water. But we could not say that the water from deep tube wells was necessarily bad because that may not have been true. (International volunteer #2, 2011)

In essence, the net social impact of the organization's activities was dependent not only on the amount of water it was selling to the villagers but also on how safe the alternative options for

drinking water were. That is, if a household shifted from another safe water source to the case organization's water, the arsenic contamination would not have decreased and there would be no social impact.

The external stakeholders also argued that water sales were an insufficient reference point for evaluating whether the organization was fulfilling its mission. Unlike for-profit companies, which use endogenous measure of performance as their reference points (see Shinkle, 2012), social enterprises have to rely on exogenous measures to assess their value creation (see Santos, 2012). In developing its activities, a social enterprise can cause negative side effects, which should also be taken into account. MNC's head of social innovation, who was not directly involved in the case organization but who managed several projects in other countries, stated:

As soon as we change something somewhere, it causes disorder compared to what was previously in order, even if we do good things. At this point, you still have side effects that can be harmful for the entire project. They can even be terrible for people who suffer from them. (MNC's head of social innovation, 2011)

In a nutshell, the goal of the case organization was not to sell water but to help people to drink safe water only. The chairman of the board, who was very keen to raise the water sales quickly, had still not lost sight of the *raison d'être* of the social business:

We are looking at the people who are still drinking the wrong/bad water. (Chairman of the board, 2014)

4.2.1 The First Compromising Account and the Water Sales Analysis

Despite the debate and acknowledged shortcomings of focusing only on water sales, things did not change overnight. While it was clear that performance evaluation and impact assessment would require further work, the employees of the organization had their hands full developing its activities. Placing focus on the tangible performance metric of water sales was helpful for some. The project director in Dhaka, for instance, was very committed to improving the organization's performance and water sales when provided a clear reference point for such a comparison. As described by one of the international volunteers:

[The project director's] mission is to increase sales, clearly. His goal is to sell as much as possible. (...) Yes, it is his priority. In my opinion, his main business is to check the sales reports every month and, afterwards, to sit with us. He had the pressure and the desire to

succeed and to show his success to [the managing director] in particular. (International volunteer #3, 2014)

Still, some indicators were needed to identify whether the changes in the water sales were due to the case organization's own activities or because of some other contextual factors. To clarify the setting, the project manager based at MNC's French headquarters proposed developing a regular report called Water Sales Analysis (Figure 2). The idea was fairly simple: the regular report would mainly focus on the development of water sales but it would also include information about key external variables beyond the organization's control, such as average temperatures, rainfall and the holy month of Ramadan. All of these were known to affect the behaviour of the villagers and, hence, the demand for water. Likewise, the water sales analysis included information about how the tap points used for water distribution were set up and located, allowing for comparisons between the sales at each tap point.

The organization had now created its first accounting artefact for the evaluation of its social performance. Initiated by stakeholder expectations, the compromising account had first included recursive iteration between the referent and the criterion to go beyond initial water sales reports, leading to the creation of a new accounting artefact, the water sales analysis (see Figure 1). Following the debates on whether an internal performance figure could be used to assess the organization's social impact (Santos, 2012), the key metrics developed by water sales analysis became the penetration rate within the villages, that is, how many of the potential households living close to the tap points were using the water. As the original assumptions and forecasts about the villagers' behaviour were inaccurate, the figures were readjusted to include as potential customers only those villagers who lived close to the tap points (within 60 metres instead of 200 metres). Also, the project manager drew on the earlier results of the health impact survey to reset the expectations of average consumption (10 litres per day and per household instead of 30 litres).

It was clear from the beginning that the water sales analysis and the associated metrics were not only imperfect but also temporary (see Stark, 2009). Apart from the amount of water sold, everything else was based on estimates: the number of potential customers, the number of customers, the average number of people per household and the average household consumption. Moreover, at this stage, the organization had no idea about how and for what purposes the villagers were using the water. Still, the penetration rate metrics served as a provisional settlement, which was an acceptable compromise for organizational groups holding

different views and thus helped the organization to progress in its activities (Kaplan and Orlikowski, 2013). Everyone acknowledged that the water sales analysis was based on available data and, hence, agreed to make do with imperfect figures. At the same time, it was evident that more detailed information was required both to identify why the consumption remained low and also to assess the social impact of the activities more precisely. The members of the organization remained puzzled, as they had assumed that once the water system was set up, the villagers would rush to buy water that was not contaminated with arsenic.

We were really concerned about avoiding queues at the tap points, as we had seen in other contexts. We were wondering how to ensure that people waste as little time as possible in the process of buying their water. (Project manager in France, 2011)

Despite the imperfection of the metrics, the considerable negative feedback on penetration rate triggered the case organization to engage in problemistic search (Cyert and March, 1963; Kuusela et al., 2017) to find solutions. It was evident that substantive-technical responses (Nason et al., forthcoming) were needed, but the organization's internal capabilities were limited. The subsequent anthropological study would be significant here.

4.3 Reference Point 3 (RF3): The Anthropological Study

Although MNC had prior experience in other developing countries, it now recognized that it did not have enough knowledge of the culture in poor rural Bangladesh. Instead, the organization had relied on its strong engineering expertise and various assumptions in developing the project. Therefore, in August 2010, the case organization started to cooperate with a local Research Center (RC) that could facilitate an anthropological approach to the local culture and its water use practices through qualitative interactions with villagers.

The first report of the RC in April 2011 provided some critical insights for the organization. The basic premises about the arsenic were confirmed, but the report also began to shed light on the complex reasons for the low water sales:

The social and health reasons for setting up a water treatment plant in [the village] remain entirely valid. The need for safe drinking water in an area where ground water is dangerously contaminated with arsenic and alternative sources inadequate or insufficient is not in doubt. (...) In spite of the need for safe water, there has been no strong demand. The reasons are many. (The first report of the RC, 2011)

One of the key reasons was related to the knowledge of the local population. As the problem of arsenic is very widespread in Bangladesh (Argos et al., 2010), the organization had assumed that the local population would be aware of the dangers of using arsenic-contaminated water and that by setting up a water plant in the region, they would be responding to an urgent need. However, there was a significant gap between the scientific views and local sentiment:

There is a discrepancy between the views of researchers in France and the awareness of the local population regarding arsenic. The latter is often very weak. (RC director, 2011)

Accordingly, it became evident that improving the organization's social impact would not happen through technical improvements but through raising awareness about the arsenic problem and the health risks posed by tube well water. Moreover, the RC report also concluded that the villagers sometimes valued social prestige over health and water quality:

People are very proud to have a deep tube well. It is a symbol, a mark of social status, and I think that sometimes one is more interested in the social status than the quality of the water. I tasted the water in a household where the family was very proud because a son who worked in the Middle East was sending money to his family. They had just built a toilet room and also a deep tube well. I tasted the water, which was disgusting. There were particles floating in it. I was told that when the technicians came to install the well, they asked the family to put thirty kilos of cow dung to cement the wall when they dig so that the sand does not fall down. And this cow dung was found in the water. (RC director, 2011)

The anthropological study led to a substantial shift in how the case organization would proceed, as it challenged its strategy and ongoing operations. Social performance was also reconceptualised at this stage. While the evaluation of the social performance had previously been based on water sales and arsenic contamination, the social performance reference point was changed to include qualitative attributes, that is the organization's activities were to be compared with the expectations, habits and culture of the local population. This implied that the focus of the board was moving from internal performance towards external impact: the voice of the beneficiaries was to be heard.

The collaboration with the RC lasted for two years and provided the organization ample sociocultural knowledge about its context. Board members considered the anthropological insights very useful, and the managing director of the joint venture, who also served as a MNC representative on the board, expressed his satisfaction with this decision:

When this kind of a project is implemented by a group of technicians and engineers, we are bound to miss a few things, so it was a very good decision to get help from the RC. (Managing director, 2012)

Yet, the responses within the lower levels of the case organization were mixed. While the social team was keen on the qualitative insights and interested in taking the villagers' perceptions into account, those with a more technical approach, including the plant management team, found the ideas distracting and inherently subjective. For the technical team, the emphasis should have been kept on the water production and distribution system, and both the management of the operation and performance evaluation should have been based on measurable and objective figures, such as water quality, sales volume and leaks. Consider the following exchange from an interview with the chemist at the water plant:

Interviewer: And if the customer says there is a smell?

Respondent: (...) I can say okay but do you have any evidence to show me?

Interviewer: (...) Smell, there is no evidence for the smell.

Respondent: If he thinks the water has a smell, he could take a sample in a small bottle and send it to me. I will then test it and determine the problem. If he is able to show me the evidence, then I will find out why the problem occurred. If he cannot show me anything, how can I prove it? (Chemist, 2014)

While the technical team preferred only objective input, the social team found it very useful to listen to the villagers' perceptions and views, as it helped them to better understand their behaviour and reactions. For the social team, having more knowledge of the beneficiaries' needs would improve the performance of the organization, as it would enable the organization to tailor its services to local needs and expectations. For example, it was at this stage that they learned that the villagers were not at all fond of how the organization was promoting its operations as bringing 'water to the poor'.

So there was a problem when it was said that it was water for the poor. The rich people said 'it's water for the poor people' and continued to drink their own water [from other sources]. So we tried to convince the managers of the case organization to change the approach and to emphasize that the water is for everyone. And it was not easy to do that. (RC director, 2014)

The mixed enthusiasm about the sociocultural viewpoints and the villagers' perceptions caused tension between the employees. The project director in Dhaka incorporated the findings of the anthropological study in his understanding of the situation:

The people are not well educated. They are very poor and have some ignorance about safe water. Some know about arsenic and some don't, so we're trying our best to disseminate the information. (Project director in Dhaka, 2011)

Yet, he would not openly support those who perceived the sociocultural approach as valuable. Instead, he shared the feelings of the technical team, who were not comfortable with the new emphasis. To some extent, they perceived that not only was their expertise being neglected but that they were being pushed aside.

[The social team] does not even have any meeting in our plant office. (...) So we are not a single team. I'm not invited to meetings even though I'm the plant manager. Not a single time did they call us. (Plant manager, 2014)

4.3.1 The Second Compromising Account and the CRM Tool

As previously mentioned, the anthropological study was considered to be very helpful at the board level as it provided knowledge that helped the organization develop its activities to meet the needs of the villagers. There were challenges on the lower levels of the organization, however, as internal tensions between the social and technical teams detracted from operations. Moreover, the sociocultural approach and its qualitative emphasis was not very useful for management purposes, particularly so for the technical team who nonetheless had the key responsibility for water production, quality and the tap points. Hence, the organization continued to use also the previous reference points concerning health impact and water sales, as these were still relevant for some employees and middle managers.

Following continued debate regarding how to evaluate the performance, the social manager of the organization took the initiative to progressively develop the CRM tool (Figure 3), which became the next accounting artefact in the organization's pursuit to develop its social performance evaluation. The CRM tool was designed to bridge the various conflicting views within the organization: it included both qualitative and quantitative information, and it also featured aspects important to both the social and technical teams. In addition to quantitative information on water consumption for each household in the area, there were also regular notes

on each household's behaviour and its members' rationales for using the water (or not). Hence, through iteration the process had now yielded the next provisional metrics, the expected amount of water consumed by each household. This was an acceptable compromise within the organization as there was in the metrics sufficient concurrent visibility for viewpoints emphasized by the different teams (see Figure 1). Moreover, the limits of the figures were acknowledged and there was discussion about how the different types of information could be used:

There were problems with correspondence. (...) For example, the water dealers did not always fill their registers seriously, so having a water metre made it possible to see the tap points with the greatest differences and to determine what data were reliable. As a result, we have information that is much more detailed but not 100% reliable. On the other hand, some other information is 100% reliable but is much less detailed. (International volunteer #3, 2014)

The introduction of the CRM tool was a step forward for the organization, as the detailed qualitative information allowed for the analysis not only of who was buying the water but also whether the customers were likely to be using water from other sources as well or if only some members of a household were accessing the safe water. Provisional metrics were no longer "outside" as an endpoint for the accounting process as they had been on the previous iterations, but they were increasingly interlinked in the process with feedback loops towards the reference point. The new metrics were gaining performativity in that they enabled interventions of employees, especially of the auxiliaries of the social team (see Dambrin and Robson, 2011; Skaerbaek and Tryggestad, 2010).

We could roughly say that the increase in sales means a positive social impact, but there are nuances. We learned—by doing this whole database on families—how many people are in each household, how much water they buy per month and so on. We learned that people were mixing different water sources, putting them at risk for arsenic contamination. Hence, the social impact is weak. We realized that in some families, only some people had access to our water—sometimes men, sometimes women. In some families, not everyone was drinking our water, so there are still some people who are at risk. In my opinion, only those who drink only our water are out of danger from arsenic contamination. This is another point that was introduced into the measurement of social impact, which I find very important. We need to know whether our consumers consume only our water or other sources as well to know if they are still at risk. (International volunteer #3, 2014)

Hence, the CRM tool allowed for the comparison of how well the organization's activities met the expectations and practices of the villagers, it continued to include information on water sales, and it also monitored whether the health awareness was improving. Still, from the perspective of the board, there was something missing. The CRM tool served well on the lower levels of the organization, but the number of details made it difficult for the board to see where the organization was heading as a whole. The feedback received was rather unclear, as both the referents and the criteria were ambiguous. Accordingly, together with the findings of the anthropological study and prior feedback from the water sales analysis, the feedback led to substantive-technical (Nason et al., forthcoming) changes in the operating practices. As a result, the board wanted to determine whether these strategic changes enhanced the organization's internal performance and social impact, and thus new reference points for social performance were again sought.

4.4 Reference Point 4 (RF4): The Key Performance Indicators

The engagement with the local RC brought the organization ample new knowledge on local practices, which then refined its operations. Subsequently, MNC's head of innovation, as one of the board members of the case organization, together with MNC's project manager in Paris, started asking for new metrics to evaluate the organization's performance. The organization needed a simple performance measurement system that would cover both its social and economic performance through a small number of indicators. For this purpose, the organization developed a new accounting artefact, which would later be referred to as 'key performance indicators' (KPIs).

A few years earlier, while the case organization was struggling to increase its water sales in the rural villages, it sought other sources of income to achieve financial sustainability more rapidly. In 2010, the organization began a 'Jar Business', the goal of which was to sell water in large quantities to firms and other organizations in Dhaka. The Jar Business had no social goals as such; hence, it differed substantially from the rural activities. In effect, it subsidized the rural activities and the social mission thereof. It is evident that such a subsidization strategy has the potential to create conflicts between a firm's social mission and its economic achievement (Pache and Santos, 2010). In this context, the role of the KPIs was to provide an operational measurement system for the staff to monitor both the economic and social performance of the organization in a single system:

And all of these new things [the Jar Business] were a bit disturbing for the comprehension and understanding of the project. So on [MNC] side, we clearly had a desire to build some key performance indicators to develop a clear vision. (...) It's significant to get a clear common vision of the project. Otherwise, some members of our team would try to achieve financial benefits, some would try to have a social impact and nobody would try to do both. (MNC's project manager in Paris, 2014)

The KPIs (Figure 4) were designed to include regular reporting, which would be used by the managers and also be presented to the board of directors. The goal was to be able to show whether the project was developing positively and to allow for comparisons over time. As there are considerable seasonal differences in weather, the follow-up on the KPIs was based on four-month periods; one of these periods was the wet season, during which the villagers drink less water. The actual indicators were designed and developed through an action research project and active collaboration between the case organization and one of the authors of this paper. The board was keen to have only a small number of indicators, which were to be both very operational and also easy to monitor on a regular basis.

Three main issues were eventually identified on the social dimension. The first indicator is about the scale of the case organization, namely, the number of people with access to tap points. This relates to the case organization's ability to invest in the expansion of the network. The second indicator concerns the penetration rate, that is, the case organization's ability to penetrate the area and to convert those with access to the service into actual customers. The third indicator concerns the regularity of consumption for each customer. It is important that the customers use the water regularly and that they do not also drink arsenic-contaminated water from alternative sources. As for the economic dimension, the self-financing rate was considered to be the key issue, because the case organization opted to depend only on its turnover to finance its activities. This relates to the chairman of the NGO's claim that a social business could fight poverty without the support of public funding or external subsidies.

As highlighted above, one of the implicit aims for developing the KPIs was to overcome the potential tensions between the social and economic logics within the organization. It was hoped that this single measurement system, including both social and economic indicators, would help to build a common vision of success and performance evaluation.

I believe that these KPIs are useful for the whole company to have a clear and unified way of understanding the business. As I mentioned, I think that, in our company, you have

people coming from companies, people coming from NGOs and people coming from development agencies, and this mix of people may work with different understandings of what this kind of a company is doing. (...) So if we get a clear common vision with those six or seven KPIs, I think we are on the right path to create a project that is understood by everybody. (Project manager in Paris, 2014)

However, this process was not straightforward, as there continued to be different expectations and assumptions about the metrics. For instance, MNC representatives who engaged with the case organization did not always agree on the purpose of the KPIs:

We decided to use these KPIs to create a kind of management tool, something to help to make decisions on where to put tap points. Because if I take the people coming from private companies, we'll try to improve the benefits of the company by saying "the fewer pipes you put in and the more people consume, the better". And the people looking for the social impact will say that "we have to reach the biggest village, even if it's far away from the network". So on one hand, we have a small debt, and on the other hand, we have a very large debt. And on one side, we have few people consuming, and on the other side, we have a lot of people consuming.. So it's very helpful to have this kind of tool to make the right decisions. (Project manager in Paris, 2014)

For the case organization's managing director, who also served on the board as a representative of MNC, the KPIs appeared however more like tools of external assessment and audit.

On one hand, I agree that people who analyze the project on a macro level can effectively try to establish indicators that help to evaluate (...) On the other hand, I'm much more skeptical about these KPIs in terms of evaluating the operating performance of the teams working in the field. (Managing director, 2014)

4.4.1 The Third Compromising Account and the KPIs

The development of the KPIs was different from the previous processes. On those occasions, the social performance reference points were set by the board and it was only later through a compromising account that an accounting artefact and corresponding metrics were developed to provide the information that allowed for the comparison of the organization's performance against the reference points. However, the KPIs were designed from the very beginning to allow for comparisons. Here, the process included recursive iterations between referents, criteria, accounting artefact and metrics while on previous occasions the accounts were developed more

on the basis of pre-set referents and criteria (see Figure 1). We maintain that such a process was possible due to the maturing of the organization, as it developed sufficient information and knowledge about its activities, operating context and beneficiaries. Moreover, the process could draw on the previous provisional settlements, in which the prior reference points regarding arsenic contamination and water sales were to an extent covered.

The KPIs provided concurrent visibility to different evaluative principles in the organization, facilitating discussion and providing also ample room for interpretation. This was not only the case with social and economic performance metrics but also with regards to the social and technical approaches, between which there had been continual tension. In designing the KPIs, the organization was able to draw on its experience and previous compromising accounts on performance evaluation. For instance, the social performance metrics do not include the amount of water the organization is selling, although this had been an important indicator for the technical team. At the same time, the water sales are an important element of the economic indicator. Similarly, although the KPIs are all in quantitative forms, it is acknowledged that the social performance indicators are dependent on the qualitative information collected in the field, which is something that the social team members had been emphasising. As such, the concurrent visibility helps to maintain the visibility of different organizational priorities, and also enhances the acceptability of the provisional settlement.

This is not to say that either the development or the use of the KPIs were without controversy and tension. Notably, the debates were not primarily about technical issues but key aspects of the organization's mission. One such discussion related to the question of whether the social performance measurement should include all the organization's activities or ignore those operations dedicated to bringing economic gain without any social ambitions. The most significant of these was the Jar Business, which had no social mission and, hence, diverted from the mission of the organization. Here again, board members from MNC did not originally have the same point of view:

I also think that, politically speaking, the Dhaka Jar Business is fully part of the project. It is true that it's only dedicated to the economic function and not dedicated to the social objective, but the project integrates both. So, the evaluation must integrate the system as a whole even if it doesn't look at social issues in Dhaka. (MNC's head of innovation, 2012)

We don't always agree on the interpretation. [Some] consider the Jar Business to be a full part of the project. I'm inclined to say it's not. (Managing director, 2012)

Eventually, it was decided that the Jar Business would be taken into account in the measurement of the economic indicator but that its sales would not be included in the measurement of coverage, penetration and regularity of consumption, i.e. the social indicators. At the outset, such a choice may appear arbitrary and consequently limit the reliability of the performance evaluation. Still, we maintain that this choice reflects the differences between financial performance and social performance reference points. While the reference points concerning financial performance are fixed, established and clear (e.g. Schillebeeckx et al., 2016), social performance reference points are malleable and their corresponding criteria ambiguous (Nason et al., forthcoming). This does not necessarily imply however that social performance reference points are less reliable and inconsistent. Instead, we argue that they need to be approached within a specific time-space context.

The feedback received from the KPIs was mixed. While the indicators show positive trends over time, the figures were far from the original aspirations of the case organization. Hence, following a suggestion from the chairman of the board, the board decided to integrate various levels of success into the KPIs, including different interim targets for each KPI. In essence, this was the introduction of multiple reference points (Joseph and Gaba, 2015) against which the indicators could be compared, which, paradoxically, also made the feedback more ambiguous. Nonetheless, while the KPIs did not provide a flawless representation of the case organization's performance, they did serve the organization in assessing its success in achieving its mission. As such, the case organization developed a social performance reference point and associated metrics. These met its needs and helped resolve tensions and mediate among different views, at least as an acceptable temporary settlement for the time being.

5. DISCUSSION

In this paper, we present a longitudinal qualitative case study to shed light on how a social venture forms reference points for social performance (Nason et al., forthcoming). For this purpose, we position our paper within prior discussions on the behavioural theory of the firm (see Shinkle, 2012) and draw on the recent contribution by Nason et al. (forthcoming), who provided a conceptual framework for how firms employ social performance reference points to evaluate their operations. Moreover, we make use of the prior accounting literature and draw on the idea of compromising accounts (Chenhall et al., 2013) to discuss how imperfect metrics

(see Busco and Quattrone, 2015) can be useful to organizations by bringing together and facilitating discussion between different groups of organizational actors (Stark, 2009).

Our findings elaborate on recent discussion regarding how social performance feedback is integrated into organizational change decisions (Nason et al., forthcoming). One of the key questions within prior literature on behavioural theory of the firm relates to how performance evaluation and feedback drives organizational decisions and engages firms in adaptation and strategic changes (e.g. Moliterno et al., 2014). In this paper we have highlighted how provisional and performative metrics can have a significant role in how organizations develop new ways to evaluate their social performance. Drawing on prior accounting literature (Chenhall et al., 2013), we develop a conceptual model (Figure 5) and discuss how compromising accounts including recursive interplay between reference point referent, criteria and a corresponding accounting artefact, form an arena for dialogue in an organization and bring together groups holding different evaluative principles. In this process of developing accounts, various groups can articulate their understandings of value, leading possibly to resolving of differences. These processes subsequently yield metrics, which can be used to evaluate the performance of the organization. Through crystallizing the debates and making the compromises visible, the metrics can serve as provisional settlements (see also Kaplan & Orlikowski, 2013), which are acceptable to different groups and as such have the potential to stabilize the compromising account. This was well evident in our case study, in which in each sub-process the crafted set of metrics was preceded with dialogue regarding how the social performance should in fact be evaluated. In addition, these provisional metrics and accounting artefacts are performative, implying that they do not just observe and reflect some organizational reality, but rather they directly intervene in the reality they are supposed to represent (see Vollmer et al., 2009). In other words, these accounting devices can shape the organization, as they affect how people talk about, work around and perceive the aspects the account depicts (Skaerbaek and Tryggestadt, 2010). Similarly, the accounting artefacts and metrics can provide authoritative signals on the purpose of the organization, and thereby have real consequences on which things are considered important and prioritized (Chenhall et al., 2013).

It is worth noting however that neither are metrics performative by definition, nor are all metrics and accounting artefacts performative in a similar way and extent. We maintain that in the context of social performance measurement accounting devices and metrics gain performativity through their interplay and linkages within the processes, as the feedback loops help the metrics

create impact and intervene in the organizational reality. This is illustrated by our case study setting, in which the performativity of the metrics increased as the organization developed new ways to evaluate its performance (see Figure 5). While at Reference Point 1 the metric is an isolated figure without connections to organizational processes and decisions, at Reference Point 4 the performance metrics are recursively linked with the corresponding referent, criteria and accounting artefact. At this stage, the metrics possess substantial performativity, as they are inherently involved in producing and enabling productive compromises within the organization. In this empirical context, the accounting metrics employed by the case organization progressed from detailed metrics to simpler and more quantified metrics. These increasingly minimalist metrics allow more space for interpretation and thus enable interventions in organizational practices (see Dambrin and Robson, 2011). At the same time, the iteration of several social performance reference points pushed the case organization's social performance, as through the rounds of compromising accounts the accounting artefacts and associated metrics become to draw on experiences and developments of prior processes. On this note, Nason et al. (forthcoming, p. 44) highlight how "firm responses to social performance feedback are likely to become dynamic inputs that update firm-identified stakeholder expectations or the level of firm social performance". We maintain that the idea of provisional and performative metrics could be helpful in exploring this iteration and also how, over time, the firm sets and adapts its social performance reference points vis-à-vis its stakeholder expectations.

Given that the social performance reference point criteria are ambiguous and the corresponding referents malleable (Nason et al., forthcoming), performative accounts are helpful as they can intervene in the organizational life by making particular things visible and providing space for interpretations, thus enabling opportunities for productive compromises between different organizational groups and evaluative principles (see Chenhall et al., 2013; Carlsson-Wall et al., 2016). The provisional metrics can thus create provisional settlements in the organization, but instead of forming a definite endpoint the resulting stabilization is likely to be only temporary (Stark, 2009). The recursive feedback loops between reference point referents, criteria and accounting artefacts help the organization to make sense of its own social performance and interpret the associated performance feedback, and thereby provide ground for organizational decisions on further action and possible adaptation (Moliterno et al., 2014). In addition, the performativity of the metrics can also trigger further debate regarding the appropriateness of the social performance reference points, possibly destabilizing the account and prompting the

organization into a search for alternative reference points and an associated process of compromising account.

Furthermore, our findings highlight how negotiations can occur between employees at different organizational levels in the process of forming a social performance reference point. We shed light on how employees at lower levels of an organization interpret and enact the changes in the social performance reference points decided by the organization's top management. The prior literature has shown how middle management and other organizational actors play significant roles in implementing and operationalizing strategic initiatives (e.g. Balogun and Johnson, 2005; Andon et al., 2007). Our case study highlights that this also holds true for social performance reference points: employees may react to such initiatives differently based on factors such as their education, prior experiences and general expectations. We also maintain that the dynamic nature of changing social performance reference points caused some confusion and internal tensions between the employees in our case organization, as the various groups tried to make sense of top management's expectations from their own respective positions (see Hyvönen et al., 2008). Hence, we posit that social performance reference points may differ from the established financial performance reference points. While Nason et al. (forthcoming) described social performance reference points as malleable, financial performance reference points have often been seen as fixed (e.g. Schillebeeckx et al., 2016). For this reason, we speculate that key organizational actors' interpretations of social performance reference points and subsequent responses may be more diverse and unexpected than responses to financial performance reference points. In this context, we maintain that it is valuable to acknowledge the relevance that imperfect but performative metrics have for organizations through their ability to enable dialogue, mediate among different interests, and provide platforms for stabilization and creation of productive compromises.

We argue that the conceptual idea of compromising accounts (Chenhall et al., 2013) is useful for understanding how different accounts and tools can enhance internal collaboration in the context of social entrepreneurship, where tensions between different forms and logics of evaluation are often present (see Nicholls, 2009; Pache and Santos, 2010). Despite their imperfection, such accounts can create productive friction by giving visibility to the various viewpoints and, hence, facilitating discussion within the organization. Given the ambiguous nature of social performance (Nason et al., forthcoming), we posit that social entrepreneurs may benefit from acknowledging that imperfect accounts are useful for managing operations, rather than using excessive resources to develop polished performance measurement systems (cf.

Molecke and Pinske, 2017). Moreover, preparing and utilizing compromising accounts may also aid social entrepreneurs to form shared reference points for social and financial performance. While Nason et al. (forthcoming) conceptualized social performance reference points as separate from those of financial performance, based on our case study, we argue that social businesses may want to consider forming shared reference points for both social and financial performance to align the different logics and to avoid mission drift.

This paper is not without limitations. Like any qualitative case study, its interpretation is based on a single organization, meaning that these findings cannot be generalized to other contexts. We maintain, however, that our analysis enhances understanding of how a particular phenomenon—in this case, the use of social performance reference points—occurs in a natural context (see Bédard and Genrdon, 2004), thereby providing insights for further studies. Furthermore, we acknowledge that our intervention in the action research project needs to be taken into account in interpreting our findings; however, our intervention was relatively marginal and mainly limited to the construction of the KPIs. Hence, we argue that the findings and conclusions presented here provide a solid basis for further enquiries.

6. CONCLUDING REMARKS

In this paper we have elaborated on how a social venture forms reference points for social performance. Drawing on prior accounting literature and the idea of compromising accounts, we show how provisional and performative metrics have a significant role in how organizations develop new ways to evaluate their social performance. Performative accounts are helpful as they can intervene in the organizational life by making particular things visible, providing space for interpretations, and facilitating discussion, thus creating temporary settlements and enabling opportunities for productive compromises between different organizational groups and evaluative principles. We argue that such performative metrics help the organization make sense of its own social performance and interpret the associated performance feedback from its stakeholders, thereby providing ground for organizational decisions on further action. By explicating on the significance of provisional and performative metrics, we hope to have opened up new ways of seeing how organizations set, formulate and interpret social performance reference points, and subsequently incorporate social performance in their strategic behaviour.

REFERENCES:

- Ahrens, T., & Chapman, C. S. (2006). Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society*, 31(8), 819–841.
- Andon, P., Baxter, J., & Chua, W. F. (2007). Accounting change as relational drifting: A field study of experiments with performance measurement. *Management Accounting Research*, 18(2), 273–308.
- Anner, J. (2016). *Blended Value Accounting and Social Enterprise Success*. Retrieved from <http://scholarworks.waldenu.edu/dissertations/1956/>
- Ansoff, I. (1979). *Strategic Management*. Macmillan: London.
- Aranda, C., Arellano, J. & Davila, A. (2017). Organizational learning in target setting. *Academy of Management Journal*, 60(3), 1189-1211.
- Argos, M., Kalra, T., Rathouz, P. J., Chen, Y., Pierce, B., Parvez, F., ... Hasan, R. (2010). Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. *The Lancet*, 376(9737), 252–258.
- Balogun, J., & Johnson, G. (2005). From intended strategies to unintended outcomes: The impact of change recipient sensemaking. *Organization Studies*, 26(11), 1573–1601.
- Barrett, C. B., & Carter, M. R. (2010). The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Applied Economic Perspectives & Policy*, 32(4), 515–548.
- Battilana, J., Sengul, M., Pache, A.-C., & Model, J. (2015). Harnessing productive tensions in hybrid organizations: The case of work integration social enterprises. *Academy of Management Journal*, 58(6), 1658–1685.
- Bebbington, J., Unerman, J., & O'Dwyer, B. (2014). *Sustainability accounting and accountability*. Routledge.
- Bédard, J., & Gendron, Y. (2004). Qualitative research on accounting: some thoughts on what occurs behind the scene. In Humphrey, C. and Lee, B. (eds.) *The Real Life Guide to Accounting Research*, pp. 191-207. Elsevier: Oxford.

- Bellucci, M., Bagnoli, L., Biggeri, M., & Rinaldi, V. (2012). Performance measurement in solidarity economy organizations: The case of fair trade shops in Italy. *Annals of Public and Cooperative Economics*, 83(1), 25–59.
- Briers, M., & Chua, W. F. (2001). The role of actor-networks and boundary objects in management accounting change: a field study of an implementation of activity-based costing. *Accounting, Organizations and Society*, 26(3), 237–269.
- Burchell, S., Clubb, C., Hopwood, A., & Hughes, J. (1980). The roles of accounting in organizations and society. *Accounting, Organizations and Society*, 5(1), 5-27.
- Busco, C. & Quattrone, P. (forthcoming). In search of the “Perfect One”: How accounting as a maieutic machine sustains inventions through generative in-tensions. *Management Accounting Research*.
- Busco, C., & Quattrone, P. (2015). Exploring how the balanced scorecard engages and unfolds: articulating the visual power of accounting inscriptions. *Contemporary Accounting Research*, 32(3), 1236–1262.
- Carlsson-Wall, M., Kraus, K. & Messner, M. (2016). Performance measurement systems and the enactment of different institutional logics: Insights from a football organization. *Management Accounting Research*, 32 (45-61).
- Carnochan, S., Samples, M., Myers, M., & Austin, M. J. (2013). Performance measurement challenges in nonprofit human service organizations. *Nonprofit and Voluntary Sector Quarterly*, 3(46), 1014-1032.
- Chenhall, R. H., Hall, M., & Smith, D. (2013). Performance measurement, modes of evaluation and the development of compromising accounts. *Accounting, Organizations and Society*, 38(4), 268–287.
- Cyert, R. M. & March, J. G. (1963). *A behavioral theory of the firm*. John Wiley: Englewood Cliffs.
- Dambrin, C. & Robson, K. (2011). Tracing performance in the pharmaceutical industry: Ambivalence, opacity and the performativity of flawed measures. *Accounting, Organizations and Society*, 36, 428-455.

- Devoto, F., Duflo, E., Dupas, P., Parienté, W., Pons, V. (2012). Happiness on tap: Piped water adoption in urban Morocco. *American Economic Journal : Economic Policy*, 4 (4), 68–99.
- Duflo, E. (2010). *Lutter contre la pauvreté : Le développement humain*. Seuil.
- Ebrahim, A., & Rangan, V. K. (2014). What Impact? A Framework for Measuring the Scale and Scope of Social Performance. *California Management Review*, 56(3), 118–141.
- Englund, H., Gerdin, J. & Burns, J. (2011). 25 Years of Giddens in accounting research: Achievements, limitations and the future. *Accounting, Organizations and Society*, 36, 494-513.
- Fiegenbaum, A., Hart, S., & Schendel, D. (1996). Strategic reference point theory. *Strategic Management Journal*, 219–235.
- Gavetti, G., Greve, H. R., Levinthal, D. A., & Ocasio, W. (2012). The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals*, 6(1), 1–40.
- Gendron, Y. (2009), Discussion of ‘the audit committee oversight process’: advocating openness in accounting research, *Contemporary Accounting Research*, 26(1), 123-134.
- Goretzki, L. & Messner, M. (2016) Coordination under uncertainty: A sensemaking perspective on cross-functional planning meetings, *Qualitative Research in Accounting & Management*, 13(1), 92-126.
- Greve, H. R. (2002). Sticky aspirations: Organizational time perspective and competitiveness. *Organization Science*, 13(1), 1–17.
- Holmes Jr, R. M., Bromiley, P., Devers, C. E., Holcomb, T. R., & McGuire, J. B. (2011). Management theory applications of prospect theory: Accomplishments, challenges, and opportunities. *Journal of Management*, 37(4), 1069-1107.
- Hu, S., Blettner, D., Bettis, RA. (2011). Adaptive aspirations: performance consequences of risk preferences at extremes and alternative reference groups. *Strategic Management Journal* 32(13): 1426–1436.
- Hyvönen, T., Järvinen, J., & Pellinen, J. (2008). A virtual integration—The management control system in a multinational enterprise. *Management Accounting Research*, 19(1), 45-61.
- Joseph, J., & Gaba, V. (2015). The fog of feedback: Ambiguity and firm responses to multiple aspiration levels. *Strategic Management Journal*, 36(13), 1960-1978.

- Kaplan, S. & Orlikowski, W. (2013). Temporal Work in Strategy Making. *Organization Science*, 24(4), 965-995.
- Kilfoyle, E., Richardson, A. J., & MacDonald, L. D. (2013). Vernacular accountings: Bridging the cognitive and the social in the analysis of employee-generated accounting systems. *Accounting, Organizations and Society*, 38(5), 382-396.
- Kraus, K., Kennergren, C. & Van Unge, M. (forthcoming). The interplay between ideological control and formal management control systems: A case study of a non-governmental organization. *Accounting, Organizations and Society*.
- Kuusela, P., Keil, T., & Maula, M. (2017). Driven by aspirations, but in what direction? Performance shortfalls, slack resources, and resource-consuming vs. resource-freeing organizational change. *Strategic Management Journal*, 38(5), 1101-1120.
- Luo, X., Wang, H., Raithel, S., & Zheng, Q. (2015). Corporate social performance, analyst stock recommendations, and firm future returns. *Strategic Management Journal*, 36(1), 123-136.
- Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative science quarterly*, 48(2), 268-305.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage: London.
- Miller, P. & Power, M. (2013). Accounting, Organizing, and Economizing: Connecting Accounting Research and Organization Theory. *The Academy of Management Annals*, 7(1), 557-605.
- Molecke, G., & Pinkse, J. (2017). Accountability for social impact: A bricolage perspective on impact measurement in social enterprises. *Journal of Business Venturing*, 32(5), 550-568.
- Moliterno, T. P., Beck, N., Beckman, C. M., & Meyer, M. (2014). Knowing your place: social performance feedback in good times and bad times. *Organization Science*, 25(6), 1684-1702.
- Mook, L., Chan, A., & Kershaw, D. (2015). Measuring Social Enterprise Value Creation. *Nonprofit Management and Leadership*, 26(2), 189-207.
- Mulgan, G. (2010). *Measuring Social Value*. Stanford Social Innovation Review.

- Mäkelä, H., Gibbon, J. & Costa, E. (2017). Social Enterprise, Accountability and Social Accounting. *Social and Environmental Accountability Journal*, 37(1), 1-5.
- Nason, R., Bacq, S., & Gras, D. (forthcoming). A Behavioral Theory of Social Performance: Social Identity and Stakeholder Expectations. *Academy of Management Review*.
- Nicholls, A. (2009). “We do good things, don’t we?”: “Blended Value Accounting” in social entrepreneurship. *Accounting, Organizations and Society*, 34(6–7), 755–769.
- Pache, A-C. & Santos, F. (2013), Inside the Hybrid Organization: Selective Coupling as a Response to Competing Institutional Logics. *Academy of Management Journal*, 56(4), 972-1001.
- Pache, A-C., & Santos, F. (2010), When Worlds Collide: The Internal Dynamics of Organizational Responses to Conflicting Institutional Demands. *Academy of Management Review*, 35(3), 455-476.
- Patton, M.Q. (2002), *Qualitative Evaluation and Research Methods*, 3rd ed., Sage, Thousand Oaks, CA.
- Perrini, F., Vurro, C., & Costanzo, L. A. (2010). A process-based view of social entrepreneurship: From opportunity identification to scaling-up social change in the case of San Patrignano. *Entrepreneurship and Regional Development*, 22(6), 515–534.
- Petkova, A. P., Wadhwa, A., Yao, X., & Jain, S. (2014). Reputation and decision making under ambiguity: a study of US venture capital firms' investments in the emerging clean energy sector. *Academy of Management Journal*, 57(2), 422-448.
- Santos, F. M. (2012). A Positive Theory of Social Entrepreneurship. *Journal of Business Ethics*, 111(3), 335–351.
- Schillebeeckx, S. J., Chaturvedi, S., George, G., & King, Z. (2016). What do I want? The effects of individual aspiration and relational capability on collaboration preferences. *Strategic Management Journal*, 37(7), 1493–1506.
- Skaerbaek, P. & Tryggestadt, K. (2010). The role of accounting devices in performing corporate strategy. *Accounting, Organizations and Society*, 35, 108-124.

Stark, D. (2009). *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton UP, NJ.

Shinkle, G. A. (2012). Organizational aspirations, reference points, and goals: Building on the past and aiming for the future. *Journal of Management*, 38(1), 415-455.

Vollmer, H., Mennicken, A. & Preda, A. (2009). Tracking the numbers: Across accounting and finance, organizations and markets. *Accounting, Organizations and Society*, 35, 108-124.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Hubbard, T. D. (2016). Reputation as a benefit and a burden? How stakeholders' organizational identification affects the role of reputation following a negative event. *Academy of Management Journal*, 59(1), 253-276.

Yin, R. K. (2009). *Case study research: Design and Methods*. SAGE publications: Thousand oaks.

Appendix A: Archival documentary material

Internal documents:

- Board Meeting Minutes
- Financial Reports & Business Plans
- Water Sales Reports & Water Sales Analysis
- CRMs and Field Work Reports

Reports produced with external partners regarding the evaluation of the case organization:

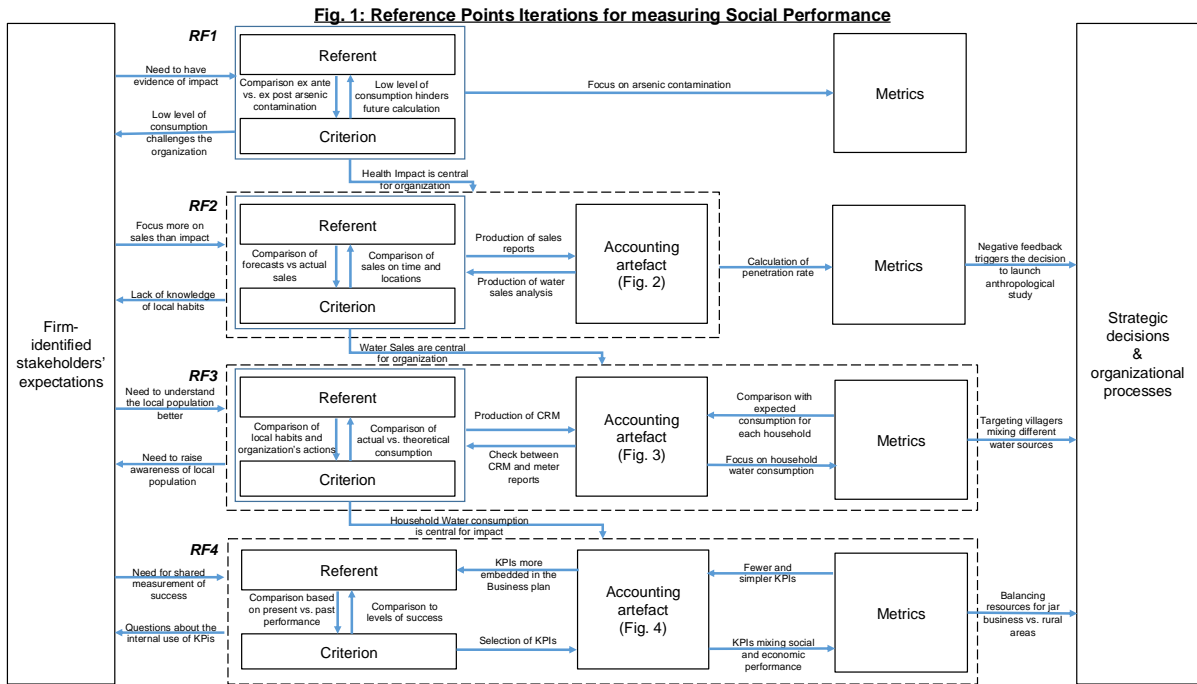
- Health impact study report produced by external research centers
- Anthropological study reports produced by external research center
- Key Performance Indicators reports produced internally and designed with the academic partner (from March 2012 onwards)

Other documents include:

- Market surveys carried out by academic partners (in May 2008 and in May 2010)
- Documents designed for external communication either through press, internet website, social media, leaflets, conferences or classes
- Documents on the management of the research action project
- Bibliography
- Miscellaneous material, such as maps, organizations charts, action plans and so forth

Appendix B: List of interviews

| Interviews | Country of staff | Number of interviews | Organization |
|--------------------------------|-------------------------|-----------------------------|---------------------|
| Project Director | Bangladesh | 3 | Joint-Venture |
| Managing Director | India | 2 | Joint-Venture |
| Chairman of the Board | Bangladesh | 1 | Joint-Venture |
| Plant Manager | Bangladesh | 1 | Joint-Venture |
| Chemist | Bangladesh | 1 | Joint-Venture |
| SocioEconomic Officer | Bangladesh | 1 | Joint-Venture |
| Board Member | Bangladesh | 2 | NGO |
| Jar Business Salesman | Bangladesh | 1 | Joint-Venture |
| Accountant | Bangladesh | 1 | Joint-Venture |
| Jar Business Marketing Manager | Bangladesh | 1 | Joint-Venture |
| Plant Supervisor | Bangladesh | 1 | Joint-Venture |
| Auxiliary #1 | Bangladesh | 1 | Joint-Venture |
| Auxiliary #2 | Bangladesh | 1 | Joint-Venture |
| Auxiliary #3 | Bangladesh | 1 | Joint-Venture |
| Auxiliary #4 | Bangladesh | 1 | Joint-Venture |
| Auxiliary #5 | Bangladesh | 1 | Joint-Venture |
| Social Manager | Bangladesh | 1 | Joint-Venture |
| CSR Manager | India | 1 | MNC |
| Head of Innovation | France | 3 | MNC |
| Project Manager | France | 3 | MNC |
| Wholesaler | Bangladesh | 1 | Joint-Venture |
| Plant Operator and Assistant | Bangladesh | 1 | Joint-Venture |
| International Volunteer #1 | Bangladesh | 1 | Joint-Venture |
| International Volunteer #2 | Bangladesh | 1 | Joint-Venture |
| International Volunteer #3 | Bangladesh | 2 | Joint-Venture |
| International Volunteer #4 | Bangladesh | 1 | Joint-Venture |



**Fig. 2: Accounting Artefact #1
Water Sales Analysis (Dec. 2011)**

| Global village penetration rate | | | | | | | | | |
|---------------------------------|------|-----------------|--------|------------|--------|-------------------------|------------|------|----------|
| Month | Days | Village (TP+PC) | | | | Global Penetration Rate | Evol 11/10 | | |
| | | Connec | L | Evol 11/10 | No. HH | | | | L/pot HH |
| janv-10 | 31 | 11 | 10 680 | | 218 | 67 580 | 16% | | |
| févr-10 | 28 | 12 | 13 978 | | 218 | 61 040 | 23% | | |
| mars-10 | 31 | 13 | 20 330 | | 218 | 67 580 | 30% | | |
| avr-10 | 30 | 13 | 21 752 | | 218 | 65 400 | 33% | | |
| mai-10 | 31 | 13 | 21 079 | | 218 | 67 580 | 31% | | |
| juin-10 | 30 | 13 | 15 691 | | 218 | 65 400 | 24% | | |
| jul-10 | 31 | 20 | 22 482 | | 368 | 114 080 | 20% | | |
| août-10 | 31 | 20 | 36 767 | | 368 | 114 080 | 32% | | |
| sept-10 | 30 | 21 | 42 759 | | 368 | 110 400 | 38,7% | | |
| oct-10 | 31 | 21 | 41 656 | | 368 | 114 080 | 36,5% | | |
| nov-10 | 30 | 21 | 35 898 | | 368 | 110 400 | 33% | | |
| déc-10 | 31 | 21 | 28 155 | | 368 | 114 080 | 25% | | |
| janv-11 | 31 | 24 | 22 688 | 212% | 368 | 114 080 | 20% | 126% | |
| févr-11 | 28 | 24 | 26 654 | 191% | 368 | 103 040 | 26% | 113% | |
| mars-11 | 31 | 24 | 36 947 | 182% | 368 | 114 080 | 32% | 108% | |
| avr-11 | 30 | 25 | 38 318 | 176% | 368 | 110 400 | 35% | 104% | |
| mai-11 | 31 | 26 | 37 911 | 180% | 368 | 114 080 | 33% | 107% | |
| juin-11 | 30 | 27 | 37 381 | 238% | 368 | 110 400 | 34% | 141% | |
| jul-11 | 31 | 27 | 38 737 | 172% | 368 | 114 080 | 34,0% | 172% | |
| août-11 | 30 | 27 | 38 130 | 104% | 368 | 110 400 | 34,5% | 107% | |
| sept-11 | 31 | 27 | 38 409 | 90% | 368 | 114 080 | 33,7% | 87% | |
| oct-11 | 30 | 28 | 40 844 | 98% | 368 | 110 400 | 37,0% | 101% | |
| nov-11 | 31 | 28 | 41 458 | 115% | 368 | 114 080 | 36,3% | 112% | |

Global Penetration rate in the village
28,5%
Mean 2010 (Jan-Dec)

Global Penetration rate in the village
31,9%
Mean 2011 (Jan-till now)

**Fig. 3: Accounting Artefact #2
CRM (March 2014)**

CRM - G1 TP 01
 Concerned auxiliary:
 Date: March 2014
 Total number of potential Households : 54 (excluding those who moved)
 Number of Households consuming in Feb: 20

Legend:
 Customers who moved to another Tap Points/village/city (05)
 Families only consuming XXX in Feb (12)
 * Mixing XXX with other source of water (08)
 * Where only some family members are consuming XXX (00)

| Zone | TP Name | Type | IDClient | Name Client | Number of migrants | Nb of people in the HH | Total size of the HH | Dec 13 - Conso In Liters | Jan 14 - Conso In Liters | Feb 14 - Conso In Liters | Amount of water that they should consume if everyone drink only XXX (cons 2L per day per adult and 1L per child) | (Now)Are they mixing XXX Water with other water | How many of them consuming XXX Water | How many of them consuming other water | Remarks |
|------|---------|------|----------|-------------|--------------------|------------------------|----------------------|--------------------------|--------------------------|--------------------------|--|---|--------------------------------------|--|---|
| G1 | G1TP01 | D | G1TP1-1 | | 2 | 5 | 7 | 310 | 310 | 280 | 300 | No | all | | Dealer. |
| G1 | G1TP01 | NC | G1TP1-2 | | 1 | 4 | 5 | | | 240 | 240 | No | all | | Previous dealer. Stopped buying XXX due to economic crisis |
| G1 | G1TP01 | C | G1TP1-3 | | 1 | 6 | 7 | 410 | 40 | 450 | 360 | No | all | | They consume only XXX |
| G1 | G1TP01 | C | G1TP1-4 | | 3 | 7 | 10 | 315 | 465 | 420 | 420 | No | all | | They consume only XXX |
| G1 | G1TP01 | C | G1TP1-5 | | 1 | 5 | 6 | 195 | 180 | 150 | 210 | No | all | | In Feb they were not at home for 1 week. So, the consumption was less |
| G1 | G1TP01 | C | G1TP1-6 | | 3 | 5 | 8 | 160 | 17 | 140 | 240 | yes | all | | They mix DTW water with XXX because they don't want to spend much money for water. |
| G1 | G1TP01 | NC | G1TP1-7 | | 5 | 2 | 7 | | | 120 | 120 | No | all | | They consume DTW water and wants XXX for free. Don't want to spend money for water. |
| G1 | G1TP01 | C | G1TP1-8 | | 1 | 3 | 4 | 40 | 430 | 90 | 180 | No | all | | In Dec and Feb they were not at home. So, the consumption is less |
| G1 | G1TP01 | C | G1TP1-9 | | 2 | 5 | 7 | 310 | 43 | 65 | 240 | yes | all | | They mix DTW water with XXX because they don't want to spend much money for water. |
| G1 | G1TP01 | C | G1TP1-10 | | 1 | 3 | 4 | 80 | 6 | 70 | 120 | yes | all | | They mix DTW water with XXX because they don't want to spend much money for water. |
| G1 | G1TP01 | C | G1TP1-11 | | 1 | 3 | 4 | 100 | 6 | 80 | 120 | No | all | | They were not at home for 2 months |
| G1 | G1TP01 | C | G1TP1-12 | | | | | | | | 120 | yes | all | | They mix DTW water with XXX because they |

**Fig. 4: Accounting Artefact #3
KPIs (Nov. 2015) – KPI #1**

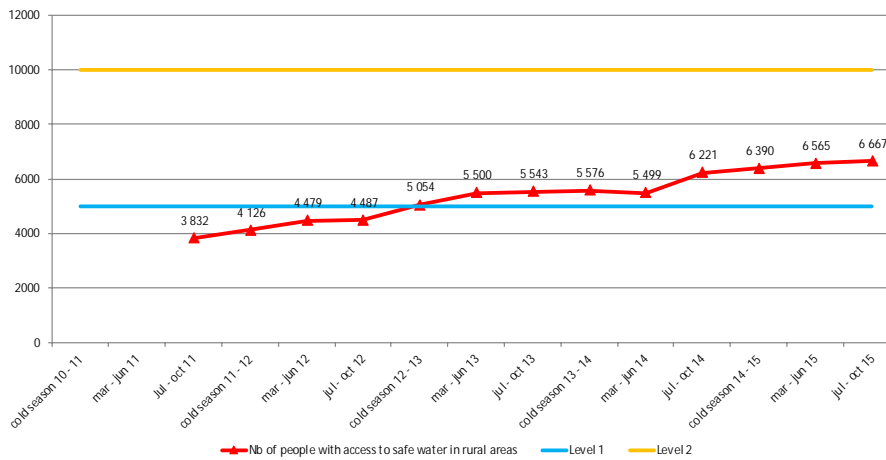


Fig. 5: Iterative process for measuring social performance

